EDUCATIONAL
RESEARCH
INSTITUTE OF
BRITISH COLUMBIA

Report No. 83:10

A STUDY OF FRENCH AND ENGLISH READING COMPREHENSION IN A GRADE 3 "PROGRAMME CADRE DE FRANCAIS" CLASS

H.C. Wood

REPORTS

REPORT NUMBER: 83:10

TITLE: A STUDY OF FRENCH AND ENGLISH READING COMPREHENSION IN A GRADE 3 "PROGRAMME CADRE DE FRANCAIS" CLASS

GRANT NUMBER: DG 331

AUTHOR(S): H.C. Wood

PROJECT DIRECTOR:
(if different from author)

PROJECT SITE OR SAMPLE: S.D. No. 23 (Central Okanagan)

This report is based on a study or project funded by the ERIBC Grants Program. This report may be reproduced in whole or in part, but the author(s) and ERIBC must be credited.

ERIBC takes no responsibility for the work as it was carried out, or the conclusions drawn by the author(s).

This report is available on a loan basis (in B.C. only) or a purchase basis at the cost of $ 4.00 from the

EDUCATIONAL RESEARCH INSTITUTE OF BRITISH COLUMBIA
Suite 701 - 601 West Broadway
Vancouver, B.C. V5Z 4C2
Telephone: (604) 873-3801

A STUDY OF FRENCH AND ENGLISH READING COMPREHENSION IN A GRADE 3 "PROGRAMME CADRE DE FRANCAIS" CLASS

by H.C. (Rick) Wood
Coordinator of French
School District #23 (Central Okanagan)

ACKNOWLEDGEMENTS

This study was completed with the financial assistance of a research grant from the Educational Research Institute of British Columbia.

Many thanks to the following persons for their assistance with this study:

M. Lynne Durwood	Research Consultant Perspectives Research Consulting Ltd.
Trish Elmslie	Secretary Typist
Dr. Shirley Henslowe	Reading Specialist
Don Marson	Assistant Superintendent School District #23
Delsa Neid	Grade 3 Teacher
Jeannine Ottman	P.C.D.F. Grade 3 Teacher
Joan Wood	Learning Assistant School District #23
Julian Wood	P.C.D.F. Student

H.C. Wood
Coordinator of French
School District #23 (Central Okanagan)

TABLE OF CONTENTS

INTRODUCTION:

Background to S.D. 23's Programme Cadre de Français (P.C.D.F.)..... 1

Background to the study..... 1

Objectives of the study..... 2

PROGRAMME FEATURES

The Programme Cadre de Français (P.C.D.F.) programme..... 4

Pupil Entry Points..... 4

Time Allotments - English and French..... 5

METHODOLOGY

Selection of pupils and classes to be involved in the study..... 6

Instruments - Canadian Cognitive Abilities Test. (C.C.A.T.)..... 9

- Gates MacGinitie Reading Tests..... 9

- Test de Lecture (French Reading Comprehension Test)..... 10

- Test Diagnostique de Lecture (French Diagnostic Reading Test)..... 11

Procedure..... 12

RESULTS OF THE STUDY

Comparison of General Intelligence - Programme Cadre de Français vs. Comparison Group (C.A.A.T.)..... 13

Confidence limits and t - values for differences between the two groups..... 14

Pupil Achievement in English: P.C.D.F. vs Comparison Group (Gates-MacGinitie)..... 15

Confidence limits and t-values for differences between Programme Cadre de Francais and Comparison Group on Gates MacGinitie Reading Tests..... 15

Variability in English achievement scores..... 18

Pupil achievement in French..... 19

Test de Lecture:

Mean raw scores and standard deviation..... 19

Raw score distribution..... 20

Test Diagnostique:

Mean raw scores and standard deviation..... 21

Raw score distribution..... 22

SUMMARY AND CONCLUSIONS 24

REFERENCES 26

APPENDICES

Appendix A: Description of t-tests and confidence limits..... 27

Appendix B: Correspondence related to the study..... 30

Appendix C: Testing Session Information..... 33

LIST OF TABLES AND FIGURES

TABLE 1 Parent Occupations.

TABLE 2 The first language and background of students in the two groups.

TABLE 3 Testing Schedule.

TABLE 4 Pupil achievement on a measure of general intelligence (C.C.A.T.): P.C.D.F. vs. Comparison group.

TABLE 5 Pupil achievement on Gates-MacGinitie Vocabulary and Comprehension Sub-tests: P.C.D.F. vs. Comparison group.

TABLE 6 Grade Equivalent range on Gates-MacGinitie: P.C.D.F. vs. Comparison group.

TABLE 7 Pupil achievement in French: Raw scores and Standard deviations.

FIGURE 1 Distribution of Raw scores on C.C.A.T.

FIGURE 2 Distribution of Raw scores on Vocabulary Sub-test - Gates-MacGinitie.

FIGURE 3 Distribution of Raw scores on the Comprehension Sub-test - Gates-MacGinitie.

FIGURE 4 Raw score distribution: Test de Lecture.

FIGURE 5 Raw score distribution: Test Diagnostique (Sentence Comprehension).

FIGURE 6 Raw score distribution: Test Diagnostique (Passage Comprehension).

A STUDY OF FRENCH AND ENGLISH READING COMPREHENSION IN A GRADE 3 PROGRAMME CADRE DE FRANCAIS (P.C.D.F.) CLASS *

INTRODUCTION

Background to School District #23's P.C.D.F.

School District #23's P.C.D.F. classes date from September, 1979, when in accordance with Ministry of Education regulations the programme was implemented in a local Kelowna elementary school. The thirty-two students who registered for the programme were grouped in two multi-grade classes, K-3 and 4-6. As the programme expanded the School District was able to arrange for single grade classes at the primary level and split classes (two grade levels) at the upper intermediate level. The P.C.D.F. is a core curriculum in French offering all the subjects usually taught in a B.C. elementary school. In this programme French is the language of instruction for 80% of the school day, except in Kindergarten, Grade 1 and Grade 2, where French is the exclusive language of instruction.

Background to the Study

This study is in part a response to a question frequently asked by parents, educators and significant others. The question, or perhaps more accurately stated, concern, focuses on whether the large proportion of instructional time devoted to French in the P.C.D.F. significantly affects the development of the pupils' reading skills in English.

Pupils in School District #23's P.C.D.F. are not introduced to formal instruction in English until the Grade 3 level at which time it occupies only 20% of the instructional day. In Kindergarten, Grade 1 and Grade 2 French is used exclusively as the language of instruction. The basic premise for this format, is that overloading in French tends to reduce interference from the first language (LI) (where LI is other than French) thus accelerating acquisition of and proficiency in, the target languate (L2).

* Programme Cadre de Français (P.C.D.F.) - Core Curriculum French

This study will determine and document the achievement of an intact class of Grade 3 P.C.D.F. students on vocabulary and passage comprehension in English using the Gates-MacGinitie reading tests. The same tests are then administered to a comparison group of Grade 3 pupils of similar ability in the regular school programme and the respective performances of the two groups are compared. The Canadian Cognitive Abilities Test was administered to both groups to determine if there were any statistically significant differences in measured performances between the two groups, in terms of their level of general intelligence. The results are presented in TABLE 4. Although of limited generalizability due to the size of the sample the study should reveal whether, along with the considerable French proficiency acquired, there is a concurrent development of English reading skills, even with the greatly reduced amount of instructional time in that language.

The study will also determine and document the level of French reading comprehension of the Grade 3 class of P.C.D.F. pupils, using the Ontario Institute for Studies in Education (O.I.S.E.) Test de Lecture and the Test Diagnostique by Margaret Tourond.

The data generated from the results obtained by students on the test instruments, will form the basis of this study and should serve as a baseline from which to measure growth in subsequent years, as well as provide information for use in making future programme decisions.

Data analysis services and assistance were provided by M. Lynne Durward of Perspectives Research and Consulting. (See Appendix B).

Objectives of the Study

The objectives of the study are:

(a) To determine and document the level of English reading comprehension and vocabulary of a class of Grade 3 P.C.D.F. pupils. Test instrument: Gates-MacGinitie.

..... 3

(b) To determine and document the level of English reading comprehension of a comparison group of pupils of similar ability but registered in the regular school programme. Test instrument: Gates-MacGinitie.

(c) To compare the respective performances in English of the two groups.

(d) To determine and document the level of French reading comprehension of the Grade 3 class of P.C.D.F. pupils using the O.I.S.E. Test de Lecture and the Test Diagnostique (Margaret Tourond).

(e) To summarise the scholastic effects of participation in P.C.D.F. Scholastic effects for the purposes of this study will be limited to:

1. Development of and achievement in English vocabulary and reading comprehension.

2. Reading comprehension in French.

..... 4

PROGRAMME FEATURES

The Programme Cadre de Francais P.C.D.F.

This programme is a French Language Core Curriculum which provides the French linguistic minority population of British Columbia with access to French Education. French is used as the exclusive language of instruction, except for the time allotted to English Language Arts.

The pupils are also expected to attain a high level of proficiency in oral and written English because the English Language Arts programme is continuously enhanced and reinforced by the effects of the English milieu.

Pupils entering the programme are expected to have some proficiency in French although lack of fluency in French does not automatically engender denial of entry into the P.C.D.F. In School District #23 the entry level proficiency in French has varied considerably.

The prescribed programme includes Français, Mathématiques, Sciences Humaines, Musique, Education Physique, Arts Plastiques, and English Language Arts. The English Language Arts Programme represents no more than 20% of the total instructional time and in the case of the particular P.C.D.F. group being studied, was not fully implemented until January, 1982. Accordingly the P.C.D.F. Grade 3 group received even less instruction in English than would normally be the case.

Pupil Entry Points

Number of pupils who commenced French in:	
Kindergarten	7
Grade 1	5
Grade 2	9
Number of Grade 3 pupils in P.C.D.F.	21

Time Allotments for English and French

The time available for formal English instruction in the P.C.D.F. out of the weekly 1500 minutes of school time is 300 minutes. This compares to the regular English Language Arts Programme of 675 minutes weekly. The difference in formal English instructional time is thus quite considerable and this should be borne in mind when examining achievement scores in English. The time allotment for English is applicable only after Grade 2 since formal English is not taught until Grade 3 in the P.C.D.F.

..... 6

METHODOLOGY

Sample - Selection of pupils and classes to be involved in the study

The sample for the study consisted of one intact class of Grade 3 P.C.D.F. students (n = 21) and one intact class of Grade 3 regular English programme students (n = 27). Practical considerations favoured the selection of the comparison group from within the school where the French programme is offered. The two classes in the study have a heterogeneous mix of students from a wide range of socio-economic backgrounds. Since most of the P.C.D.F. students are bussed in to the one school where the programme is offered, the possibility of having two groups with completely dissimilar socio-economic backgrounds is high. Accordingly the occupational endeavours of the parents of the two groups were tabulated to provide information on that one aspect of student background. Parental occupations are listed in TABLE 1.

..... 7

TABLE 1

Parent Occupation

P.C.D.F. Group		Comparison Group	
Homemaker	11	Homemaker	9
Sales	4	Unemployed	5
Unemployed	3	Receptionist	2
Teacher	3	Sales	2
Manager	2	Shop Owner	2
Secretary	1	Clerk	1
Mechanic	1	Insurance	1
Biologist	1	Hairdresser	1
Dentist	1	Chamermaid	1
Construction	1	Groundsman	1
Carpenter	1	Parks Department	1
Agent	1	Bricklayer	1
Administrator	1	Vintner	1
Doctor	1	Trailer Builder	1
Nurse	1	Parts Man	1
Instructor	1	Plumber	1
		City Worker	1
		Miner	1
		Mechanic	1
		Lab Technician	1
		Rodeo Rider	1
		Truck Driver	1
		Repairman	1
		Manager	1

TABLE 2

The first language and backgrounds of students in the two groups

P.C.D.F. (n = 21)				Comparison Group (n = 27)			
1st Language		Language spoken at home		1st Language		Language spoken at home	
English	17	English	13	English	23	English	23
French	4	French	8	French	1	French	1
Other	-	Other	-	Other	3	Other	3

The above table indicates that there is some reinforcement of the French language taking place at home in the P.C.D.F. group.

The Canadian Cognitive Abilities Test was administered to both groups and the results analysed to determine whether there was any statistically significant difference between the Programme Cadre group and the Comparison group on any of the measures.

INSTRUMENTS

Canadian Cognitive Abilities Test

The Canadian Cognitive Abilities Test C.C.A.T. Primary Battery, published by Nelson Canada (1982) is designed to assess the development of cognitive abilities related to verbal, quantitative, and non-verbal reasoning and problem solving. The Primary Battery C.C.A.T. test is composed of four sub-tests: Relational Concepts, Object Classification (Multimental), Quantitative Concepts and Oral Vocabulary. This test is administered in no fewer than two sittings. The scores that an individual obtains on the test as a whole will reflect the extent to which the following skills and competencies have been developed:

1. Ability to comprehend oral English.
2. Ability to follow directions.
3. Ability to hold material in short-term memory.
4. Possession of effective strategies for scanning pictorial and figural stimuli to obtain either specific or general information.
5. Possession of a store of general information and verbal concepts.
6. Ability to compare stimuli and detect differences in relative size, position, quantity, shape and time.
7. Ability to classify, categorize or order familiar objects.
8. Ability to use quantitative and spatial relationships and concepts.

The respective achievement scores obtained by the two groups on the C.C.A.T. were compared to ascertain primarily whether there were any statistically significant cognitive differences between them.

Gates-MacGinitie Reading Tests (Level C)

The Gates-MacGinitie Reading test published by Nelson Canada (1980), is a standardized reading test widely used in school districts and designed to indicate how well students in that class tested, score on a standard reading task, compared to a nationwide representative group of children. Level C consists of a Vocabulary Test and a Comprehension Test.

Vocabulary

This test samples the child's reading vocabulary. It contains forty-five items. Each item consists of a test word followed by four words or phrases. The child's task is to choose the one word or phrase that is closest in meaning to the test word. The first items are composed of easy and commonly used words. The later words are less common and more difficult.

Comprehension

The Comprehension Test measures the child's ability to read complete prose passages with understanding. It contains twenty-two different passages with two questions about each passage. The passages represent various subject matter fields. The questions include those that require an understanding of information that is explicitly stated in the passage, and those that require an understanding of information that is only implicit in the passage.

Test de Lecture (French Reading Comprehension Test)

The Test de Lecture was developed by Barik and Swain and published by the Ontario Institute for Studies in Education (1979). The test consists of a number of passages followed by questions based on their content. The passages were either adapted from already published material or written by a specialist in children's literature. The passages vary in length and type of material treated. The questions tap the pupil's comprehension of the contents of the passages and inferences that he or she can make from them. The Grade 3 level test consists of nine passages and twenty-eight questions. The tests are designed to assess reading comprehension rather than the mechanics of reading. Each level of the test went through several experimental versions in the developmental phase and was then revised on the basis of item analysis.

Test Diagnostique de Lecture (French Reading Comprehension Test)

The Test Diagnostique was designed by Margaret Tourond and published by the Ontario Institute for Studies in Education (1982). The test for Grade 3 consists of two parts. Part 1 is vocabulary and sentence comprehension and Part 2 is story comprehension.

Part 1 Compréhension de Phrases

Part 1 consists of thirty-three questions which test vocabulary in the context of a sentence. The questions test the pupil's ability to deduce the meaning of an unknown word with the context of a sentence. They also provide information regarding vocabulary acquired and actual decoding difficulties.

Part 2 Compréhension de Textes

Part 2 consists of twenty-four questions based on short passages in French. The questions are designed to verify the pupil's comprehension of the various texts and determine the inferences that can be made from them.

PROCEDURE

The timetable of testing activities is presented in TABLE 3.

TABLE 3

Testing Schedule

	P.C.D.F. Group	Comparison Group
Monday June 7th	O.I.S.E. Test de Lecture	Gates-MacGinitie
Tuesday June 8th	Test Diagnostique	C.C.A.T.
Wednesday June 9th	Gates-MacGinitie	
Thursday June 10th	C.C.A.T.	

Dr. Shirley Henslowe supervised the administration of the tests to ensure that testing conditions were as similar as possible for both groups and that the same procedures were adopted during the testing period. Testing session information is provided in Appendix C. J. Ottman and H.C. Wood administered and supervised the French testing.

RESULTS OF THE STUDY

Comparison of General Intelligence - P.C.D.F. vs. Comparison Group

There were no statistically significant differences in pupil achievement between the P.C.D.F. class and the comparison group on the C.C.A.T. measure of general intelligence. The class means and distribution of raw scores obtained on the C.C.A.T. are presented in TABLE 4 and FIGURE 1 respectively. Confidence limits and t-values are also contained in TABLE 4. A detailed description and explanation of confidence limits and t-tests is contained in Appendix A.

TABLE 4

Pupil achievement on measure of General Intelligence P.C.D.F. vs Comparison Group

Results on the Canadian Cognitive Abilities Test C.C.A.T.

	P.C.D.F. (n = 20)		Comparison (n = 27)		Diff. Betw. Means	Upper Conf. Limit	Lower Conf. Limit	* t-value
	Mean	S.D.	Mean	S.D.				
C.C.A.T. Total Score	73.7	10.2	75.0	8.4	-1.3	4.2	-6.8	0.49 ns

S.D. = Standard Deviation

ns = no statistically significant difference between means

.05 level of significance for confidence limits

Note regarding t-tests: one assumption of the t-test for independent samples is that variances are equal. A test of homogenity of variance was run before each t-test to ensure that variances of the samples were not statistically different. They were not, so Perspectives Research proceeded with the t-tests.

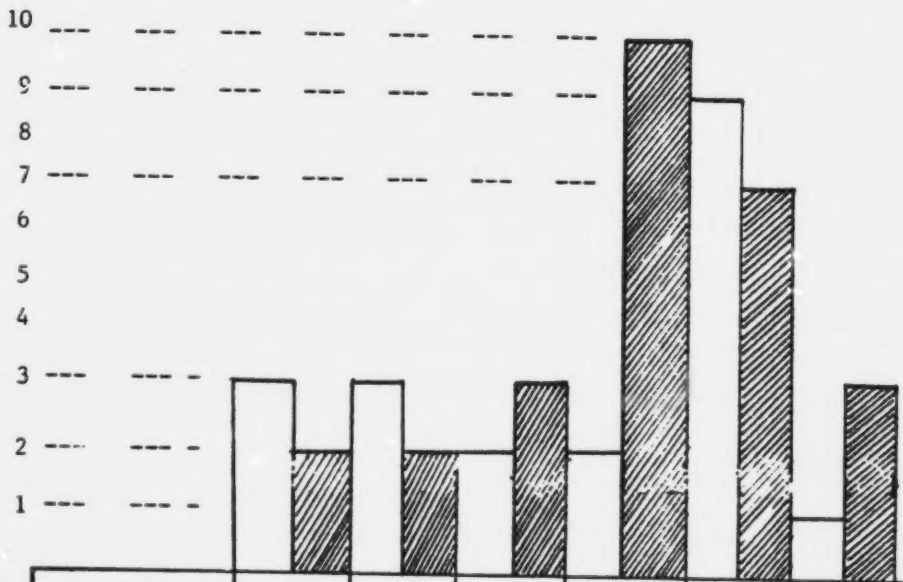
* See Appendix A. Explanation of t-tests and confidence limits.

FIGURE 1

Distribution of Raw Scores on Measure of General Intelligence (Canadian Cognitive Abilities Test)

P.C.D.F. vs. Comparison Group

(Maximum 90)



Raw Scores	55- 60	61 - 66	67 -72	73 - 78	79-84	85- 90	Total
P.C.D.F.	3	3	2	2	9	1	20
Comparison Group	2	2	3	10	7	3	27

..... 15

Pupil Achievement in English: P.C.D.F. vs Comparison Group (Gates-MacGinitie)

The results obtained by the two classes on the vocabulary and comprehension sub tests of the Gates-MacGinitie English reading tests are presented in TABLE 5. Figures 2 and 3 provide more detailed information on the distribution of raw scores obtained by pupils on the vocabulary and comprehension measures. Confidence limits and t-values are also included in TABLE 5. A detailed description and explanation of confidence limits and t-tests is contained in Appendix A.

TABLE 5

Results on the Gates-MacGinitie Vocabulary and Comprehension Sub-tests

P.C.D.F. vs Comparison Group

Gates-MacGinitie Sub-test	P.C.D.F. (n = 20)		Comparison (n = 27)		Diff. Betw. Means	Upper Conf. Limit	Lower Conf. Limit	t-value
Vocabulary	Mean	S.D.	Mean	S.D.				
(Max = 45)	31.1	11.0	33.4	9.4	-2.3	3.7	-8.3	0.78 ns
Comprehension (Max = 44)	30.6	10.8	29.9	9.8	0.7	6.8	-5.3	0.24 ns

S.D. = Standard Deviation

ns = no statistically significant difference between means

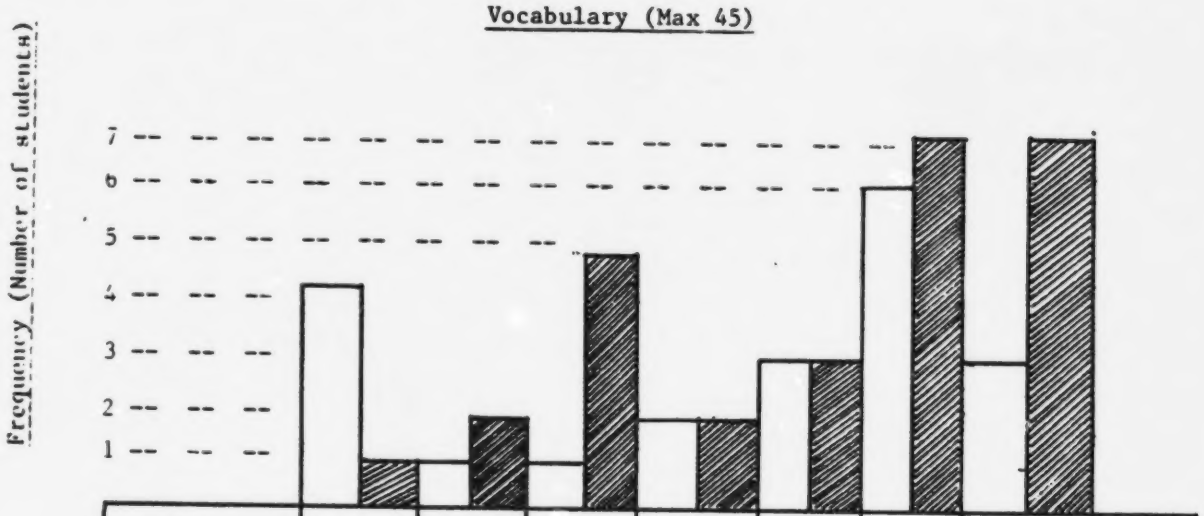
0.5 level of significance for confidence limits

..... 16

FIGURE 2

Distribution of raw scores obtained by the two groups on the Gates-MacGinitie Vocabulary Sub-test

P.C.D.F. vs. Comparison Group

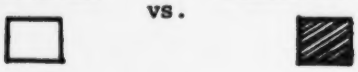


Raw Score	12 - 16	17 - 21	22 - 26	27 - 31	32 - 35	37 - 41	42 - 45	Total
P.C.D.F.	4	1	1	2	3	6	3	20
Comparison Group	1	2	5	2	3	7	7	27

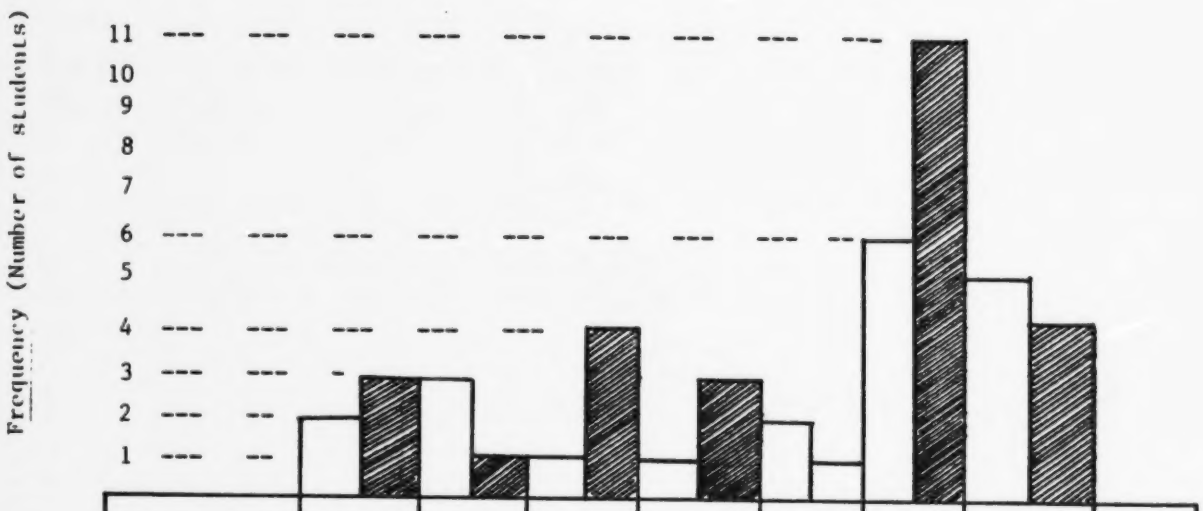
FIGURE 3

Distribution of raw scores obtained by the two groups on the Gates-MacGinitie Comprehension Sub-test

P.C.D.F. vs. Comparison Group



Comprehension (Max 44)



Raw Score	9 - 13	14 - 18	19 - 23	24 - 28	29 - 33	34 - 38	39 - 44	Total
P.C.D.F.	2	3	1	1	2	6	5	20
Comparison Group	3	1	4	3	1	11	4	27

..... 18

Variability in English Achievement Scores

Although there were no statistically significant differences in the mean class scores obtained on the Gates-MacGinitie reading tests (TABLE 5) there was a fairly wide range of achievement in both classes. In the P.C.D.F. group this could be attributed partly to the considerably reduced instructional time in English and partly due to the fact that some of the students are non-native speakers of English. It is interesting to note however, that the P.C.D.F. class as a whole scored slightly higher than the comparison group on the English comprehension section of the Gates-MacGinitie test despite the very small amount of English instruction given to them in the class. The range of Grade Equivalents for the two groups derived from Raw scores obtained on the Gates-MacGinitie is provided in TABLE 6. The test was administered at Grade 3.9

TABLE 6

Grade Equivalent Range

	Vocabulary	Comprehension
P.C.D.F.	2.1 - 7.1	1.7 - 7.2
Comparison Group	1.8 - 7.1	1.7 - 7.2

Out of the twenty P.C.D.F. students who took the test only six were reading below Grade 3 level and four of these were non-native speakers of English. Of the remainder most were reading considerably above their Grade level in English. The results obtained by the P.C.D.F. group indicate that the level of English reading skills is certainly not below that of a comparable class in a regular English programme receiving considerably more instruction in English. This finding is consistent with results obtained in studies by Edwards and Smythe (1976), Genesee (1976) and Swain (1978).

Pupil Achievement in French

Mean Raw Scores and Standard Deviations on the French measures are presented in TABLE 7. A more detailed breakdown of individual pupil scores obtained on the two reading assessment instruments is provided in FIGURES 4, 5, and 6.

TABLE 7

Mean Raw Scores and Standard Deviations on French Measures

Programme Cadre de Francais Group

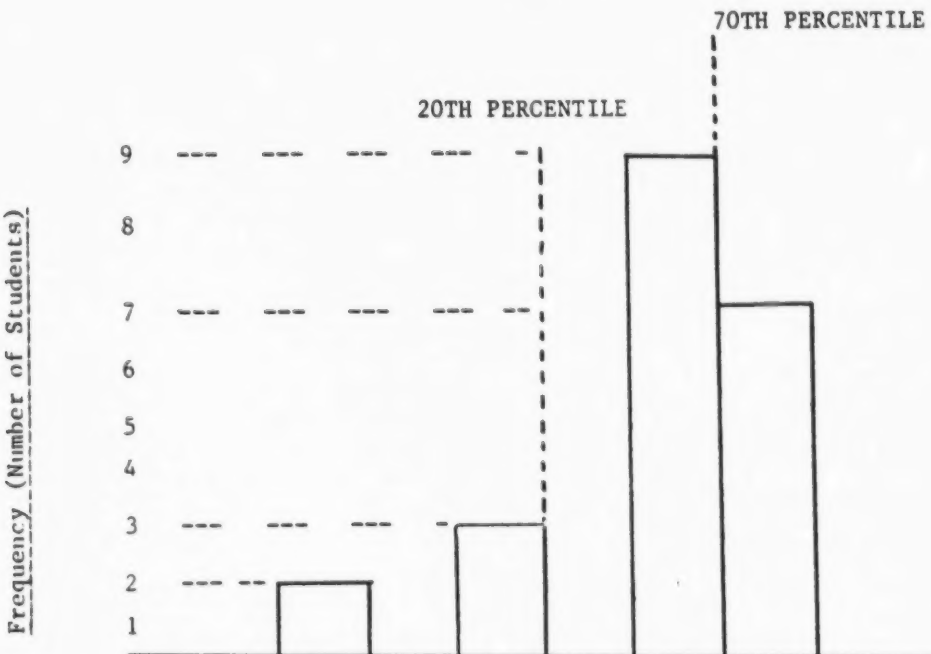
(n=21)

Measure	Mean Raw Score	Standard Deviation
O.I.S.E. Test de Lecture (Maximum = 28)	20.1	6.4
Test Diagnostique		
Comp. de Phrases (Maximum = 33)	23.3	5.6
Comp. de Textes (Maximum = 24)	15.4	4.6

FIGURE 4

RAW SCORE DISTRIBUTION

O.I.S.E. Test de Lecture (Maximum 28)



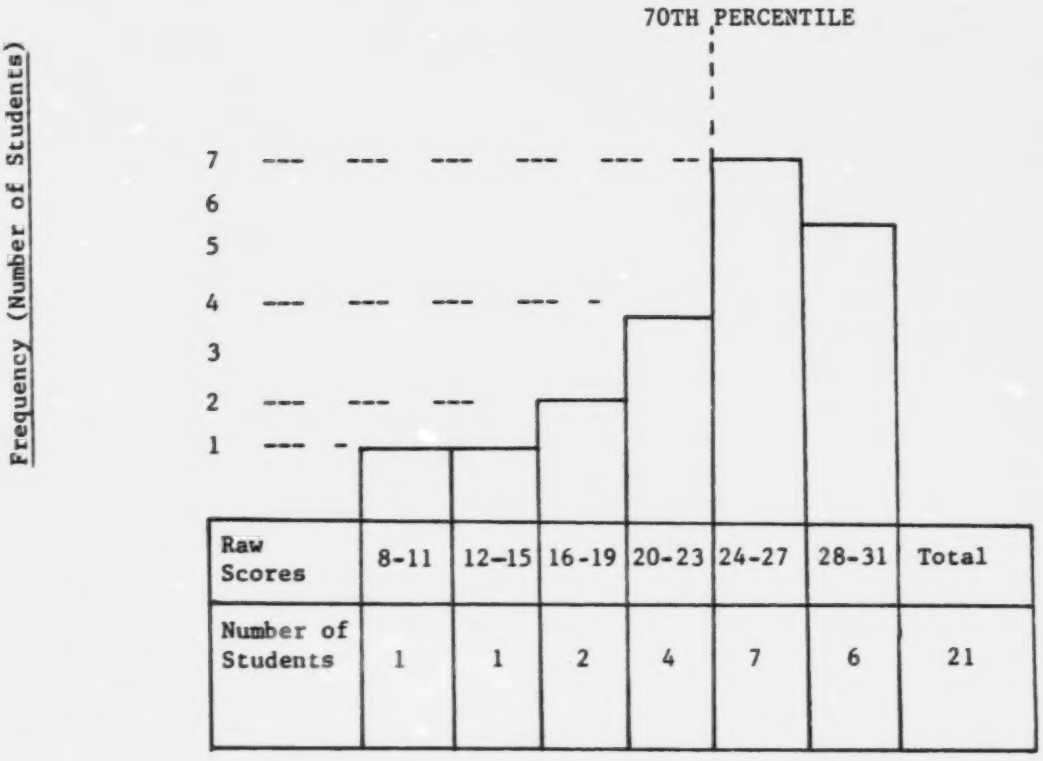
Raw Scores	4 - 7	8 - 11	12-15	16-19	20-23	24-27	Total
Number of Students	2	-	3	-	9	7	21

FIGURE 5

RAW SCORE DISTRIBUTION

Test Diagnostique (Maximum 33)

Comprehension de Phrases (Sentence Comprehension)



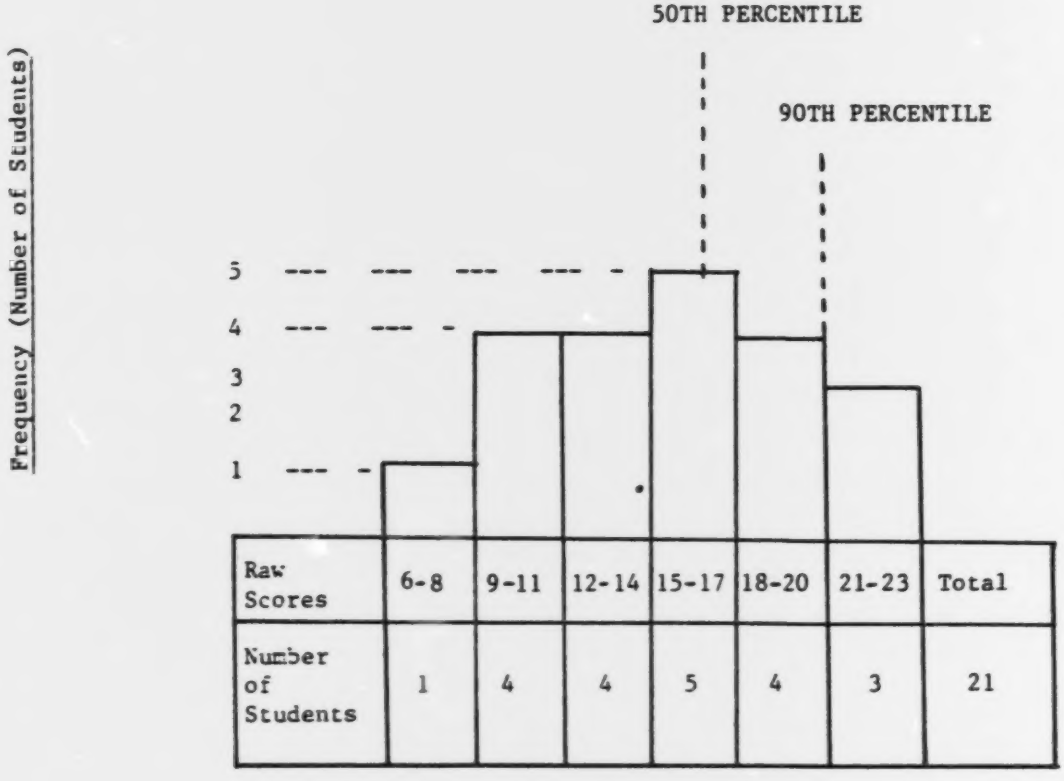
Raw Scores	8-11	12-15	16-19	20-23	24-27	28-31	Total
Number of Students	1	1	2	4	7	6	21

FIGURE 6

RAW SCORE DISTRIBUTION

Test Diagnostique (Maximum 24)

Comprehension de Textes (Passage Comprehension)



Raw Scores	6-8	9-11	12-14	15-17	18-20	21-23	Total
Number of Students	1	4	4	5	4	3	21

Pupil Achievement in French

The results obtained by the P.C.D.F. class on the two French reading measures indicate a very satisfactory level of achievement. On the Sentence Comprehension Sub-test of the Test Diagnostique, 13 of the 21 students, scored above the 70th percentile. * The French reading scores obtained by students in this study are documentations of the P.C.D.F. group's overall level of French reading comprehension. They serve also as measures of the students' ability to function entirely in French.

* See FIGURE 5

SUMMARY AND CONCLUSIONS

Programme Cadre de Français(P.C.D.F.)is now a well established educational programme in the Central Okanagan School District. Because a large proportion of the P.C.D.F. instructional day (80%) is conducted entirely in French, this evaluation study sought to determine whether, along with the considerable gains in French comprehension and communicative proficiency, there was a concurrent development of English reading skills. This study presents a series of evaluative snapshots of the P.C.D.F. and although of limited generalizability due to the small size of the sample, there is sufficient evidence in the study to support the view that at this early grade level (Grade 3), there are no measurably adverse effects on the scholastic achievement of participating students, which are directly attributable to the programme. In the area of passage comprehension in English the P.C.D.F. class performed slightly better, than the comparison class from the regular English programme, although not significantly so.

Overall the range of scores on the English reading comprehension tests is fairly typical for a Grade 3 class, in spite of the fact that the students received very little formal instruction in English. Factors which may have contributed to this performance include the ability of some students to read English prior to entry into P.C.D.F. and the extra reading in English done by students in the home environment at the urging of parents anxious that their children not lose their facility to read in English! Parents relax! In the Central Okanagan School District the scores obtained by P.C.D.F. pupils at the Grade 3 level indicate that their English language reading skill development is comparable to that of regular English programme students.

The scores obtained by students on the French reading measures (Test de Lecture and Test Diagnostique) indicate very satisfactory levels of achievement.

The top ten reading scores in French on both measures were obtained by students for whom French was not the first language learned. This finding supports, to some extent, the contention that there is considerable transfer of reading skills from one language to another. At the Grade tested, participants in the P.C.D.F. are evidently performing in both French and English at a highly acceptable level.

In summation the P.C.D.F. is an attractive French language alternative to the regular B.C. English language programme and as constituted in School District #23, is succeeding in developing and maintaining high levels of useable French and English language reading skills.

REFERENCES

Barik, H.C. and Swain, M.

Three year evaluation of a large scale early Grade French Immersion Programme: The Ottawa Study. Language Learning 1975

Edwards, H.P.

Evaluation of the French Immersion Programme offered by the Ottawa Roman Catholic Separate School Board. The Canadian Modern Language Review 1976 33

Stern, H.H.

Optimal Age: Myth or Reality. The Canadian Modern Language Review 1976 32

Swain, M.

French Immersion: Early, Late or Partial: The Canadian Modern Language Review 1978 34

APPENDIX A

Calculating the Confidence Limits for the Difference Between Two Means From Unmatched Groups

Suppose that you have in hand the results of an achievement test administered to two groups—a group (the E-group) which received a new experimental program and a control group (the C-group) which received the usual program. You have calculated the means for each group and found the difference between the two means:

E-group mean = 80

C-group mean = 70

difference between means = 10

The E-group did 10 points better than the C-group. Now you will want to know how much confidence you can place in a difference of 10 points: *If you repeated the program next year with similar conditions and similar students, would you be likely to get the same difference between the means?* How sure can you be, in other words, that this result is showing you a real difference caused by something in the programs, and not a mere unrepeatable happenstance?

The procedure described by Worksheet 3A will enable you to say something like this:

> The difference we could expect if we repeated the evaluation study might be as small as 3 points or it could be as large as 17 points; it will almost certainly be somewhere within the range: 10 ± 7.[10]

Confidence limits not only give you a good idea of the range of likely outcomes, they also provide a test of the statistical significance of the obtained difference between two groups. To see why this is so, consider the situation in which the lower confidence limit is negative. Suppose, for example, that instead of 10 ± 7, the confidence limits you obtained had been 10 ± 12. The *upper* confidence limit would be an outcome in which the E-group scored better than the C-group by 22 points; but the *lower* confidence limit would be −2, a potential difference between group means in which the C-group actually does better than the E-group by 2 points! The important thing to notice in this case is that *the confidence limits span zero.* This means that the difference between the means might, with a certain selected probability, have turned out to be zero. Whenever the confidence limits include zero within their range, there is *no* statistically significant difference between the means. In these cases, *chance* may be responsible for the particular difference between the groups obtained on this one test administration. Next time the *C-group* might score higher.

When To Use Confidence Limits

When reporting the difference between means, use either a t-test such as the one described by Worksheet 3A to determine statistical significance, *or* use confidence limits. Confidence limits are preferable in most cases, particularly if you plan to discuss the *size* of the difference between the means—nearly always of interest in program evaluation where your concern will be not only *whether* but *how much* difference a program or treatment has made. Besides, confidence limits usually are easier to explain to lay audiences than are statistics that involve calculating a number foreign to the data at hand, such as a t-value or a Mann-Whitney U.

If the E-group did better than the C-group, you should calculate confidence limits to see whether you could expect similar results in the future. If at first glance the C-group seems to have done better than the E-group, then it could be that the experimental program not only failed to improve things, but might have even caused harm. Before making this conclusion, calculate the confidence limits. Perhaps the confidence limits will include zero, indicating that zero, or *no difference*, was equally as likely as the result you got. In this case, the calculation of confidence limits shows that the lower E-group score was probably just due to chance, not a significant one.

Note that although the terms E-group and C-group are used in this discussion and in the worksheet, you might be comparing two experimental programs or two groups defined by something other than the program they have received, for instance, sex, age, or having had some experience. Remember, as well, that significance as judged by calculating confidence limits does not tell you whether or not an observed difference is *important.* You or your evaluation audience will have to judge this for yourselves by asking if it is large enough to be considered important educationally.

10. Read *10 ± 7* as *10 plus or minus 7.* Ten plus seven gives 17, the *upper confidence limit.* Ten minus seven gives 3, the *lower confidence limit* for this difference between the means.

APPENDIX A

not formed by random assignment--a test for significant pretest differences is essential. Conclusions about the final effects of a program will be strengthened if a t-test of the difference between E- and C-group *pretest means* shows *no* statistical significance. This indicates the E- and C-groups probably started out equivalent in achievement, attitudes, or whatever.

You should compute a t-test to check if the difference in *posttest* scores between two groups, usually an E- and a C-group, is statistically significant.

The t-test has non-design uses as well, all of them situations where you want to know if score differences between two groups on some measure are significant. You might want to test, for example, whether boys and girls are achieving equally well in a certain reading program. A t-test will tell if the boys' mean score is significantly different from the girls' mean score. You can use a t-test to examine the difference between *attitudes* of certain parent groups or between *program implementation* practices at different sites.

In general you can use a t-test to search out statistically significant differences between any two groups you can identify on any measure you can administer--though how you *interpret* the results will differ from one situation to another. There is one qualification to this sweeping statement, however: the t-test is most appropriate for determining the significance of the difference between means *when the number of participants in each of the two groups is about equal.* If the group sizes are quite *un*equal (say if one group is more than 20% larger than the other), then look at the standard deviation associated with each group's mean before using the t-test. If the standard deviations are similar, go ahead with the t-test. If they are quite different, you should probably use the Mann-Whitney U Test (Worksheets 3C, D, and E) instead of or in addition to the t-test. Alternatively, you could make the numbers per group equal by randomly removing scores from the larger group so that it equals the size of the smaller group, and then performing the t-test with only the data from these equalized groups.

Note, as well, that the t-test does not tell you whether or not a statistically significant difference is an *important* difference. You or your evaluation audience will have to judge this for yourselves by examining differences and asking if they are large enough to be considered important educationally.

APPENDIX A

Testing the Statistical Significance of the Difference Between Two Means: The t-test for Unmatched Groups

The t-test is a test to see if there *is* a statistically significant difference between the mean scores of two groups--say, an experimental (E) group and a control (C) group. Demonstrating whether or not a difference is statistically significant is important: a statistical test tells you how frequently your result would be expected to occur simply by chance if indeed there were no real difference in E-group and C-group performance. A difference that a statistical test determines to be *not* significant must be considered too small and chancy to be taken seriously. Some of the logic underlying the *t-test* of significance is explained in the following paragraphs.

Suppose a group of students comprising a class have all been taught in the same way all year. You arrive in April and randomly divide the class into two subgroups, giving both the same test. You would not expect to find that the scores of the two random subgroups are very different. On the other hand, the mean scores of these subgroups are not likely to be *exactly* the same either. Because all scores are susceptible to errors and variability, any two sets of test scores--even from essentially the same group--will have slightly different means. Just *how* different the two means turn out to be will depend upon:

1. The sizes of the subgroups. The larger the number in each subgroup, the more you can expect the mean of each subgroup to be the same as the mean of the whole original group.
2. The variability of the scores. The wider the variation you find among the scores, the more likely it is that the means will be, by chance selection, quite a bit different.

The t-test is designed to help you take into account these two factors--group size and score variability--when interpreting the difference you have observed between groups. If a t-test were applied in the situation just described, you would expect it to show that, given the variability of scores in the two groups, the difference between means was *not* big enough to reach statistical significance. You would conclude that the two subgroups were not really different.

Now, suppose that a group of students has been divided randomly into two groups. One has been taught by what you have been told is a good method, and the other group has been taught by a method that you suspect to be much poorer. Again, you give a test, and find the means for the two groups. Sure enough, as expected, one group has a higher mean score than the other. But you have to consider the possibility that this difference was due to chance--that the two groups are in reality performing equivalently. Only by first ruling out this possibility will you be able to consider the difference in results worth mentioning.

One way to see if the difference is too large to be just a chance fluke would be to pool all the scores from both groups and keep selecting random subgroups and recording the difference between the means. If the differences between pairs of groups obtained in this way were smaller than the difference found when you divided the students according to how they were taught, then you would conclude that teaching method had really made a difference. This procedure would work well, but it would be very time-consuming.

The t-test is a quick way of accomplishing the same end by applying what amounts to the same procedure. It answers the question: Is the *obtained* difference between the means bigger than the differences you would expect to obtain if the two groups were actually equivalent? In other words, is the difference you obtained bigger than the differences which could be expected to occur by chance sampling variation?

To apply a t-test to the difference between means, you calculate an *obtained t-value* by inserting into a formula the obtained difference between means and its associated standard deviation, representing the variability of scores. You then check the obtained t-value against a *tabled t-value*. The tabled t-value is read from a table organized according to the number of cases in each group. If the obtained t-value is larger than the tabled t-value, this means the obtained difference between means is larger than would be expected if the groups were not really different.

When To Use the t-test

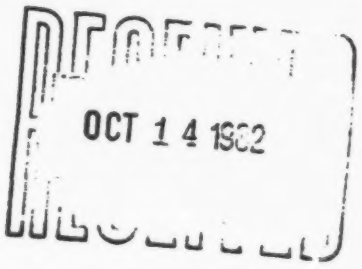
The t-test is most often used in conjunction with research and evaluation designs to scrutinize differences in scores--achievement, attitude or whatever--between experimental and control groups. You might want to use a t-test to check if *pretest* scores of two randomly composed groups are equivalent, that is, as an indicator of whether randomization has worked. The two groups can be considered equivalent if the obtained t-value is *less* than the tabled t-value. This indicates *absence* of a statistically significant difference on the measure used. A true, randomly selected, control group will almost always turn out to be equivalent to the randomly selected experimental group. In the case where you are using a *non-equivalent control group*--one

TEL. (604) 731-7616

Perspectives RESEARCH AND CONSULTING LTD.

201-C - 2515 Burrard Street,
Vancouver, B.C.
V6J 3J6

October 7, 1982



Mr. H.C. Wood
Coordinator of French
Curriculum Centre
599 Harvey Ave.
Kelowna, B.C.
V1Y 6C8

Dear Rick:

Enclosed please find results of the data analyses incorporated into three tables.

You will note (quite happily, I suspect) that there were no statistically significant differences between the Programme Cadre and the comparison group on any of the measures.

Note regarding the t-tests: one assumption of the t-test for independent samples is that variances are equal. A test of homogenity of variance was run before each t-test to assure that the variances of the samples were not statistically different (they weren't, so we proceeded with the t-test).

Also enclosed are an invoice for data analysis services and copies of a description of confidence limits and t-tests which you may find useful.

I would be interested in receiving a copy of your report with the interpretation of the results when it is available.

Best regards,

M. Lynne Durward
Research Consultant

MLD:lm
encl.

APPENDIX B

School District No. 23
(CENTRAL OKANAGAN)
CURRICULUM CENTRE

Coordinators & Consultants
Phone: 763-2202

599 Harvey Avenue
KELOWNA, B.C.
V1Y 6C8

September 22, 1982

Ms. Lynne Durward
Perspectives Research &
Consulting
201-C 2515 Burrard St.
VANCOUVER, B.C. V6J 3J6

Dear Lynne,

Enclosed please find the sets of figures compiled from the results of our Programme Cadre project.

Please proceed with the following data analysis services as outlined in your letter of September 7th, 1982.

a. calculation of means and standard deviations of student scores on the following measures:
   - OISE Test de lecture (Prog. Cadre only)
   - OISE Test Diagnostique (2 subtests - Prog. Cadre only)
   - Gates-MacGinitie (2 subtests - Prog. Cadre & Comparison group)
   - CCAT (Prog. Cadre & Comparison group)

b. t-tests and calculation of confidence limits for differences between Programme Cadre and Comparison group means on the following measures:
   - Gates-MacGinitie (2 subtests)
   - CCAT total scores

c. Preparation and typing of summary tables

If the CCAT scores of the two groups differ significantly, please inform me of the possible implications statistically and financially! I have included conversion tables from the CCAT examiners manual for your perusal.

Sincerely yours,

Rick.

H.C. WOOD
Coordinator of French

HCW:ns

Encl.

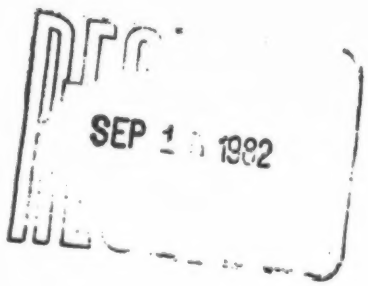
APPENDIX B

TEL. (604) 731-7616

Perspectives RESEARCH AND CONSULTING LTD.

201-C - 2515 Burrard Street,
Vancouver, B.C.
V6J 3J6

September 7, 1982



Mr. H.C. Wood
Coordinator of French
School District No. 23
Curriculum Centre
599 Harvey Ave.
Kelowna, B.C.
V1Y 6C8

Dear Mr. Wood:

As discussed during our meeting of September 3, we are prepared to offer the following data analysis services in connection with the Programme Cadre project:

a) calculation of means and standard deviations of student scores on the following measures:

- OISE Test de Lecture (Prog. Cadre only)
- OISE Test Diagnostique (2 subtests - Prog. Cadre only)
- Gates-MacGinitie (2 subtests - Prog. Cadre & Comparison group)
- CCAT (Prog. Cadre & Comparison group).

b) t-tests and calculation of confidence limits for differences between Programme Cadre and Comparison group means on the following measures:

- Gates-MacGinitie (2 subtests)
- CCAT total scores.

c) Preparation and typing of summary tables.

The estimated cost of the above services is $135. (Note: additional statistical consulting time may be necessary if it is found that the CCAT scores of the two groups differ).

I hope that we can be of assistance to you on the project and look forward to your reply.

Sincerely yours,

Lynne Durward

W. Lynne Durward
Research Consultant

MLD:lm

APPENDIX C

TESTING SESSIONS: FRENCH RESEARCH PROJECT

Test Preparation: General

1. Enough booklets for each class < C.C.A.T.
   Gates-MacGinitie (GM)
2. I.D. on booklets < Teacher
   Students
3. Class list prepared for each class.
4. Copy of test description, 1 for each teacher.

Test Administration/Equipment

1. Shirley Henslowe will administer tests.
2. Need extra pencils, erasers.
3. No interruptions (DO NOT DISTURB sign).
4. Positive attitude towards testing encouraged.
5. Teacher can introduce Shirley Henslowe to students.
6. Teacher can make sure students have reading, quiet work to do at desks after testing.
7. Students absent for 1 Sub-test can do other tests.
8. Teacher need not be present. Shirley Henslowe will send student for you in staff room.

Estimated Time including Instructions

G-M	Vocab	35 mins
	Comp	45 mins
C.C.A.T.		
	Test 1,2	40 mins
	Test 3,4	35 mins

Schedule (SH)

D. Neid	Class 2	J. Ottman	Class 1
Monday		Wednesday	
G.M.	9:45-10:25	G.M.	Same
Tuesday		Thursday	
	9:45-10:25		Same
C.C.A.T.	10:40-11:30	C.C.A.T.	Same